PRESENTED BY RED HAT

LEARN. NETWORK.
EXPERIENCE OPEN SOURCE.

www.theredhatsummit.com

OVERVIEW & ROADMAP OF VIRTUALIZATION IN RED HAT ENTERPRISE LINUX

Chris Wright
Principal Software Engineer, Red Hat
May 4, 2011

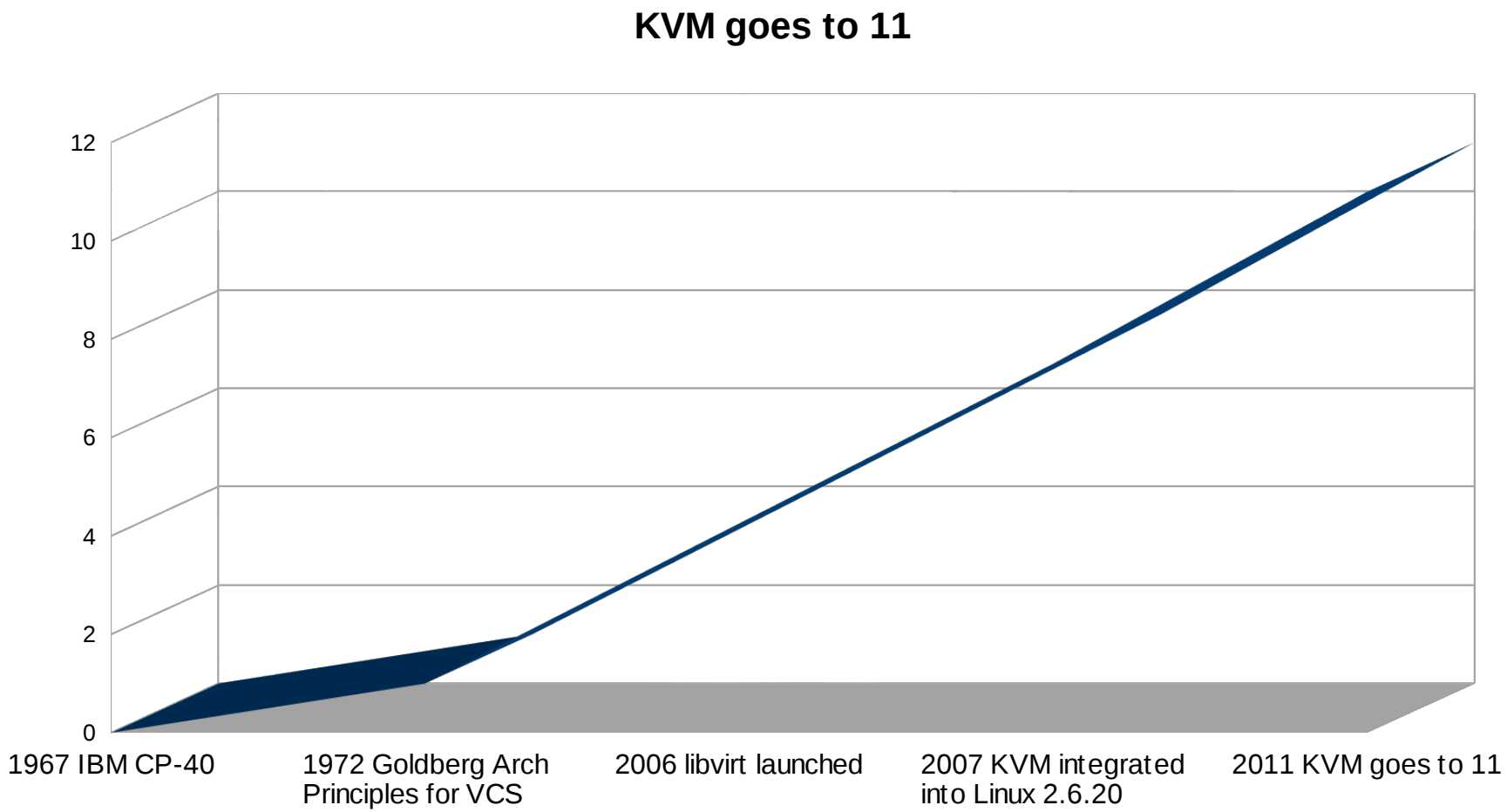
Agenda

Virtualization Overview

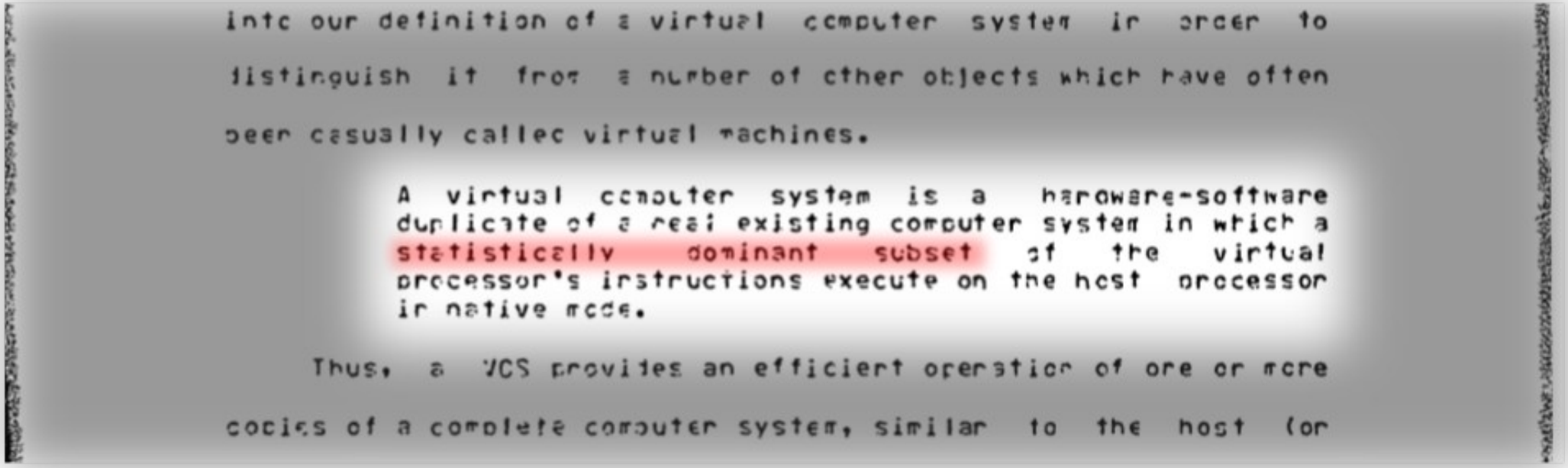
Foundation for Cloud

Roadmap

Virtualization Overview: History



Virtualization Overview: History



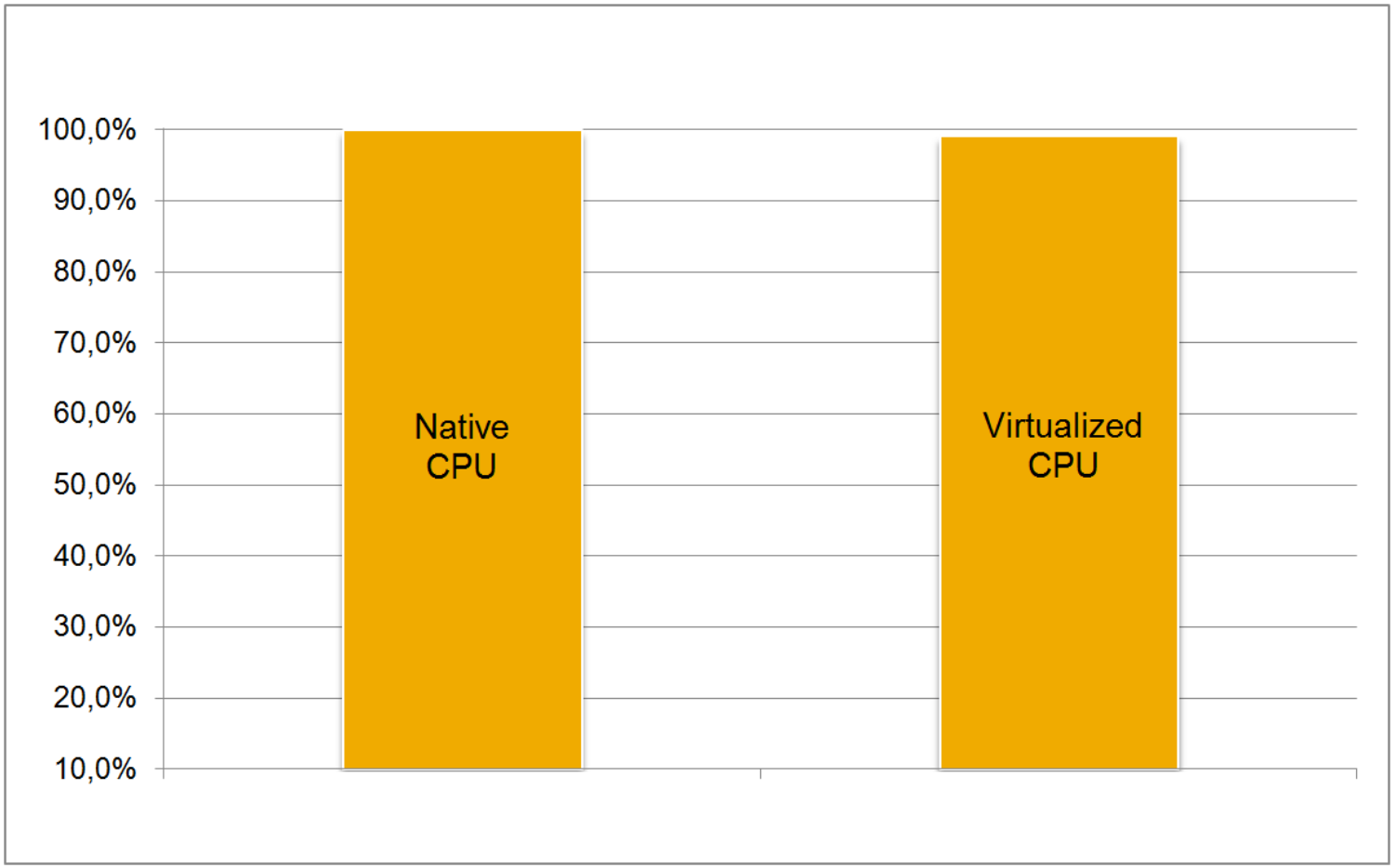
into our definition of a virtual computer system in order to distinguish it from a number of other objects which have often been casually called virtual machines.

> A virtual computer system is a hardware-software duplicate of a real existing computer system in which a statistically dominant subset of the virtual processor's instructions execute on the host processor in native mode.

Thus, a VCS provides an efficient operation of one or more copies of a complete computer system, similar to the host (or

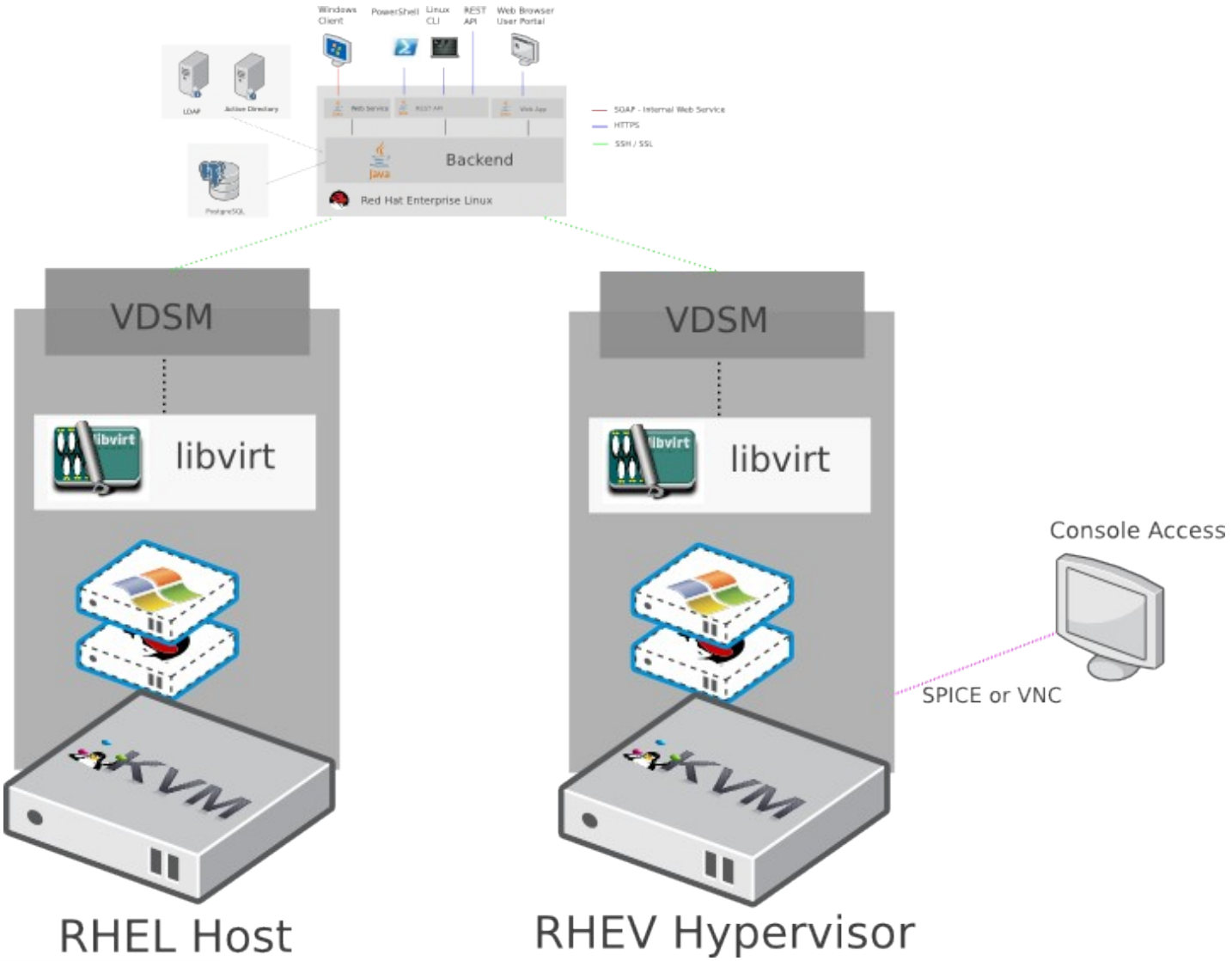
Virtualization Overview: KVM Today

Native vs. Virtualized: 1% Overhead in KVM

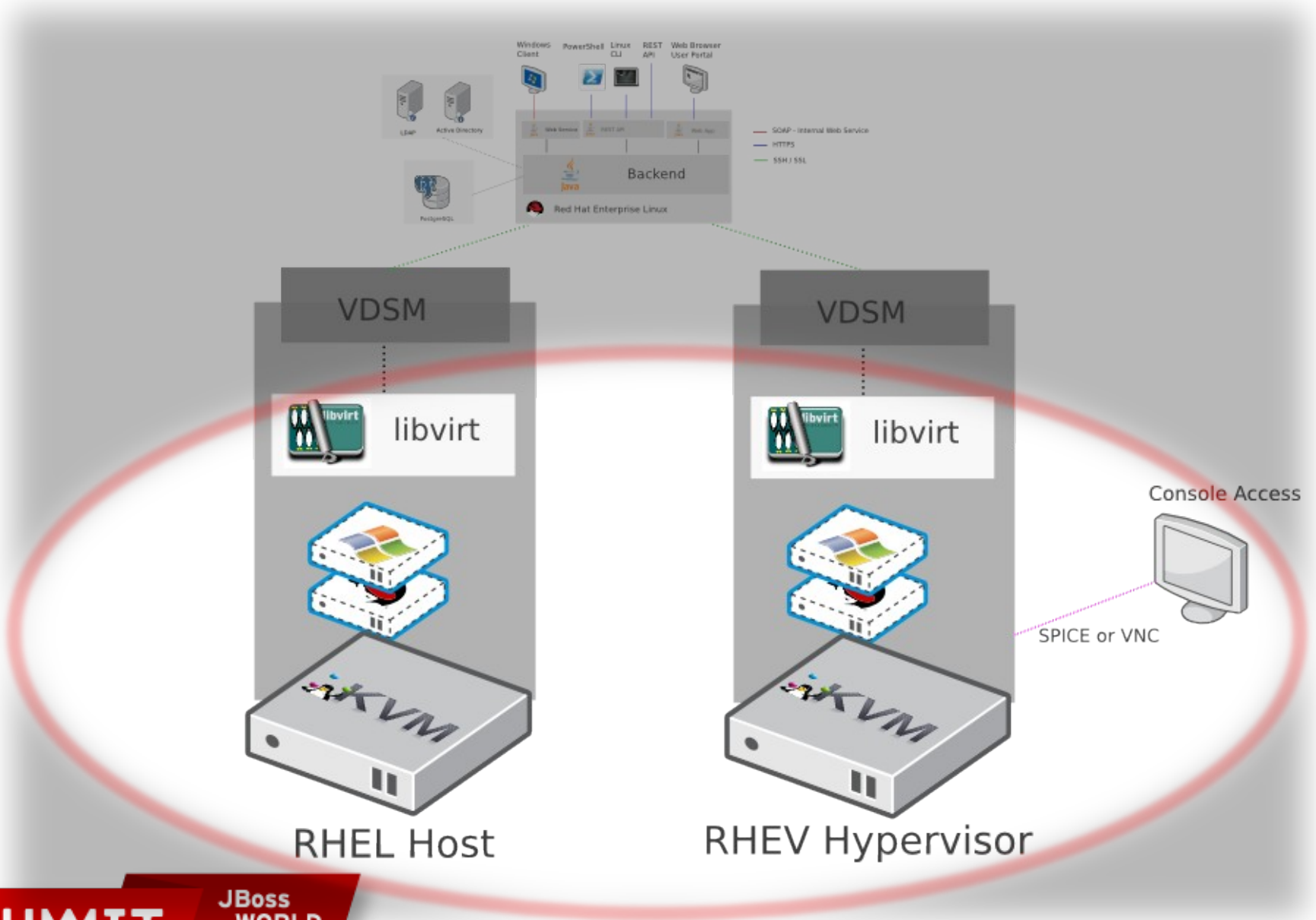


SAP Linux Certification Suite
CPU intensive reports phase
RHEL 6 + KVM and Intel E7-8800

Virtualization Overview: Architecture

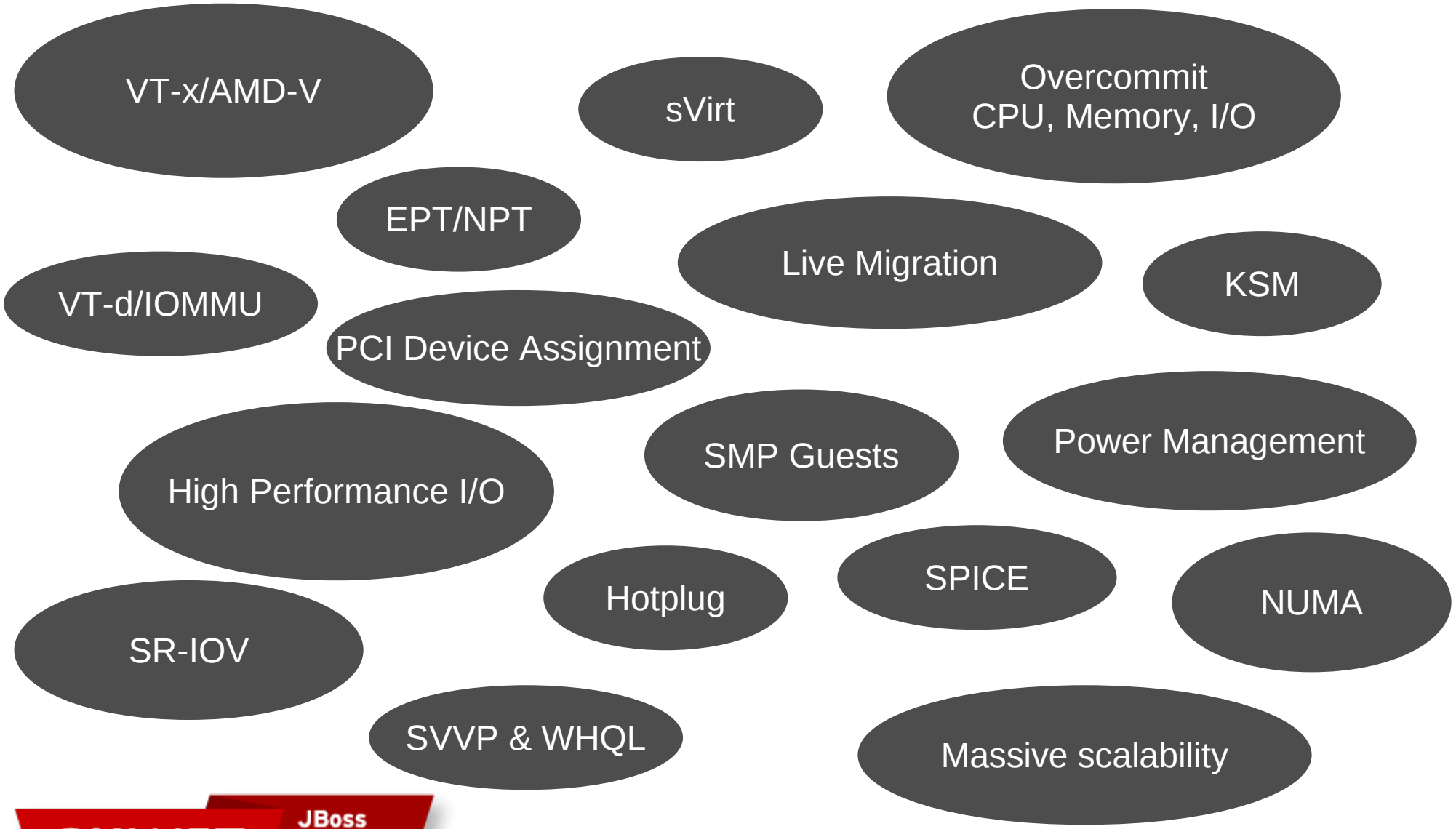


Virtualization Overview: Architecture



KVM: The Kernel-based Virtual Machine

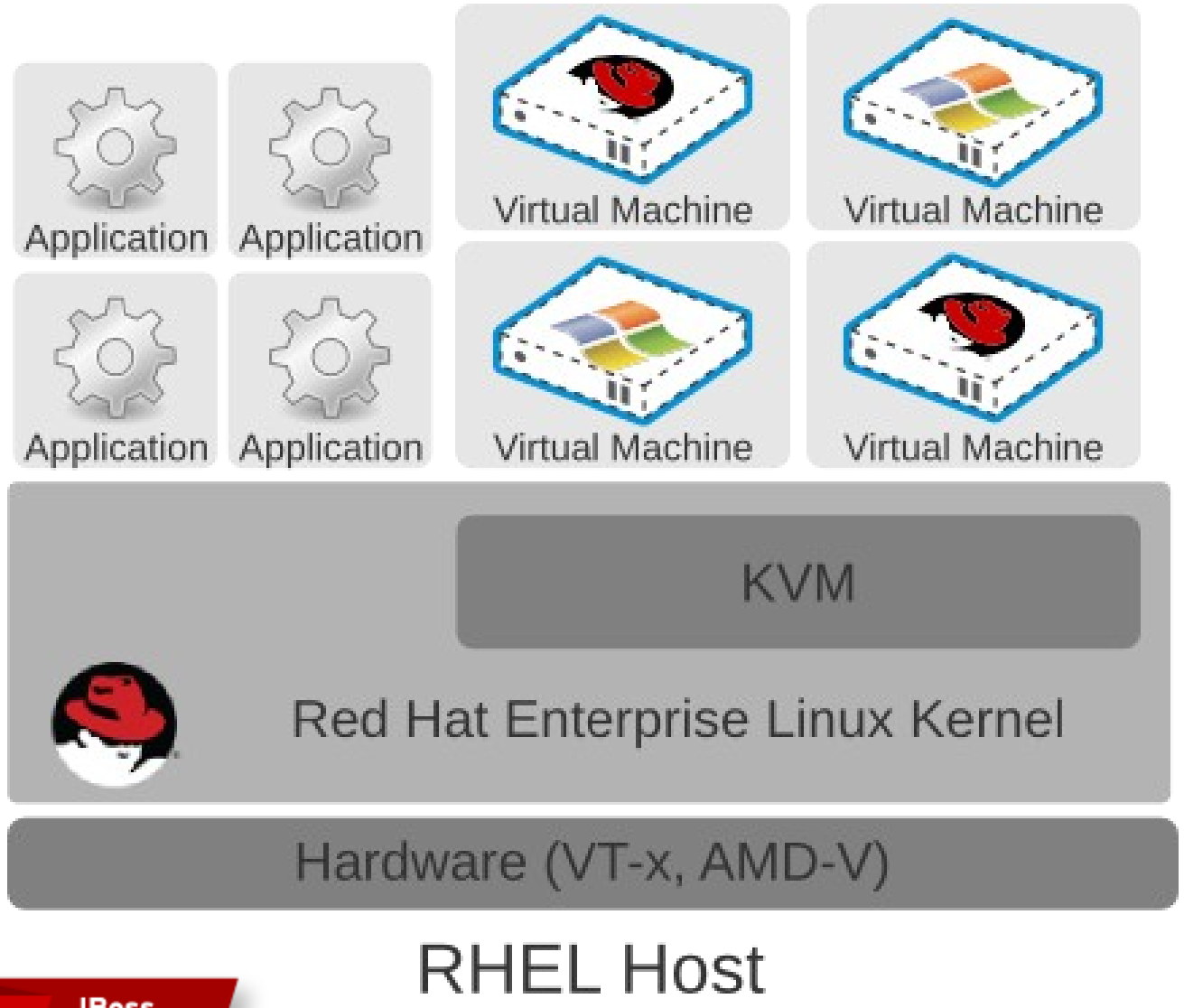
KVM features



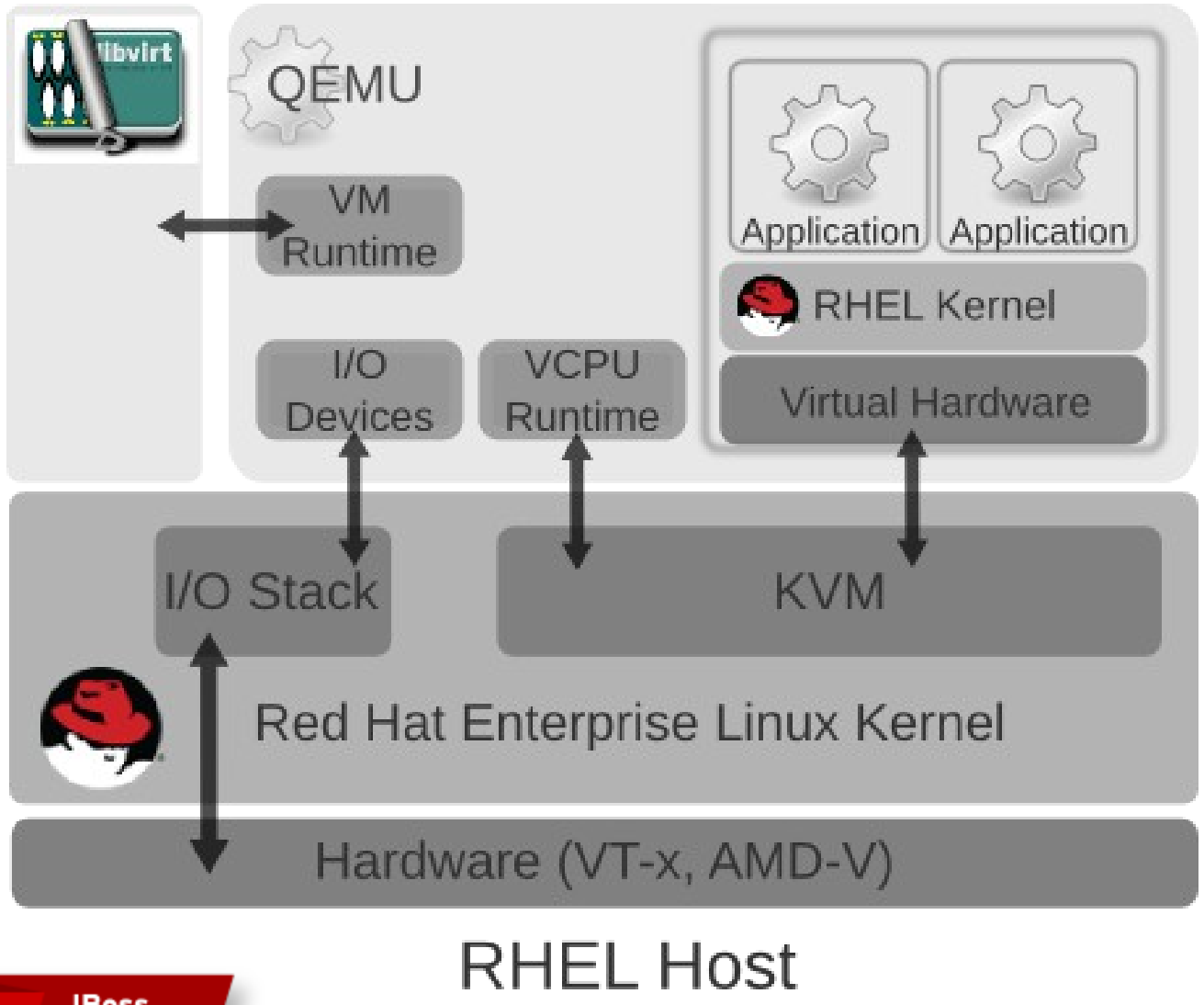
Why **reinvent** the wheel?

Focus on virtualization.

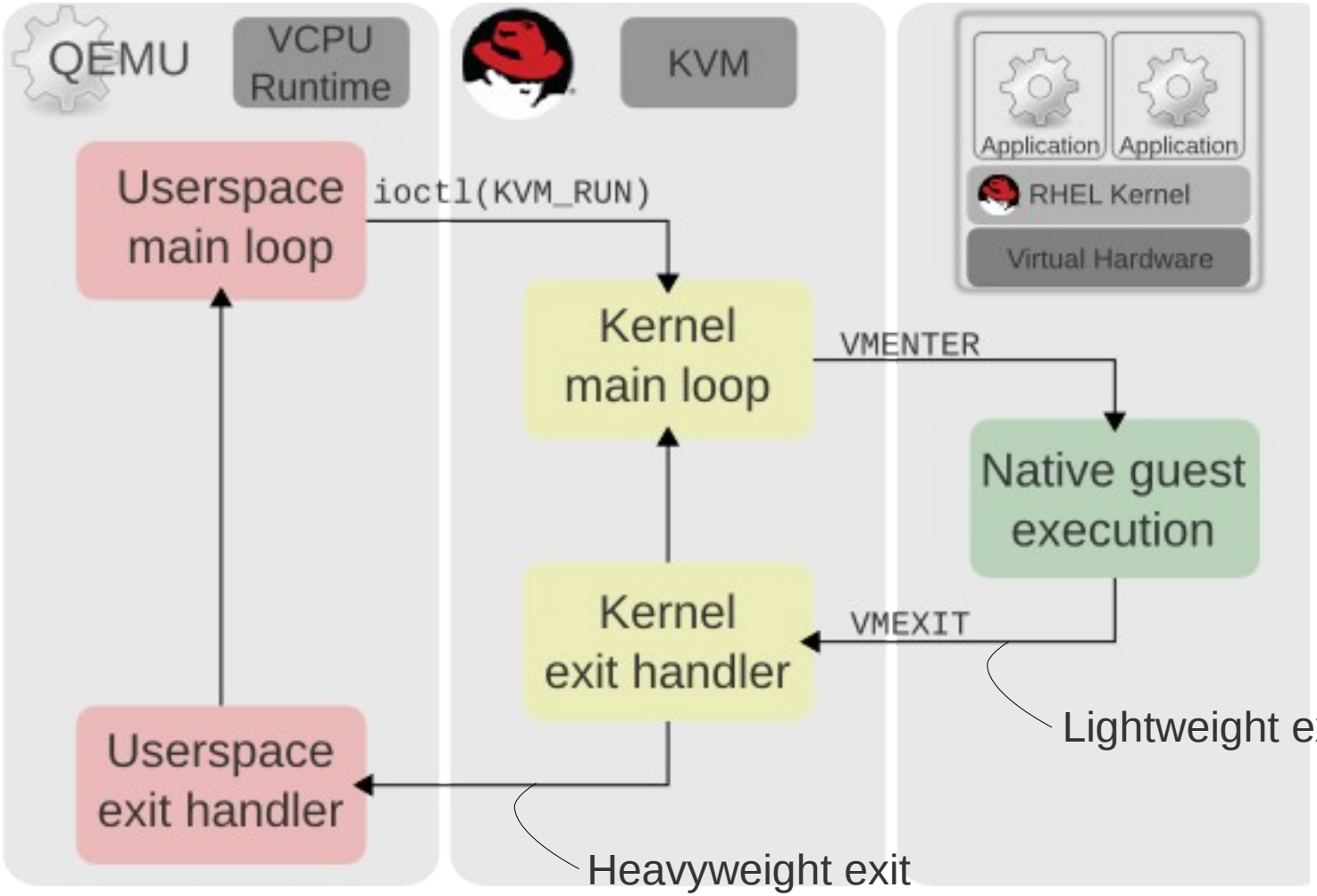
KVM Architecture



KVM Architecture



KVM Architecture



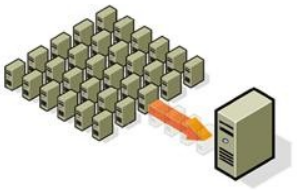
KVM in RHEL 6

- RAS
- Security Isolation
- Resource management
- Performance improvements
  - CPU, Block, Net, Memory
- SPECvirt

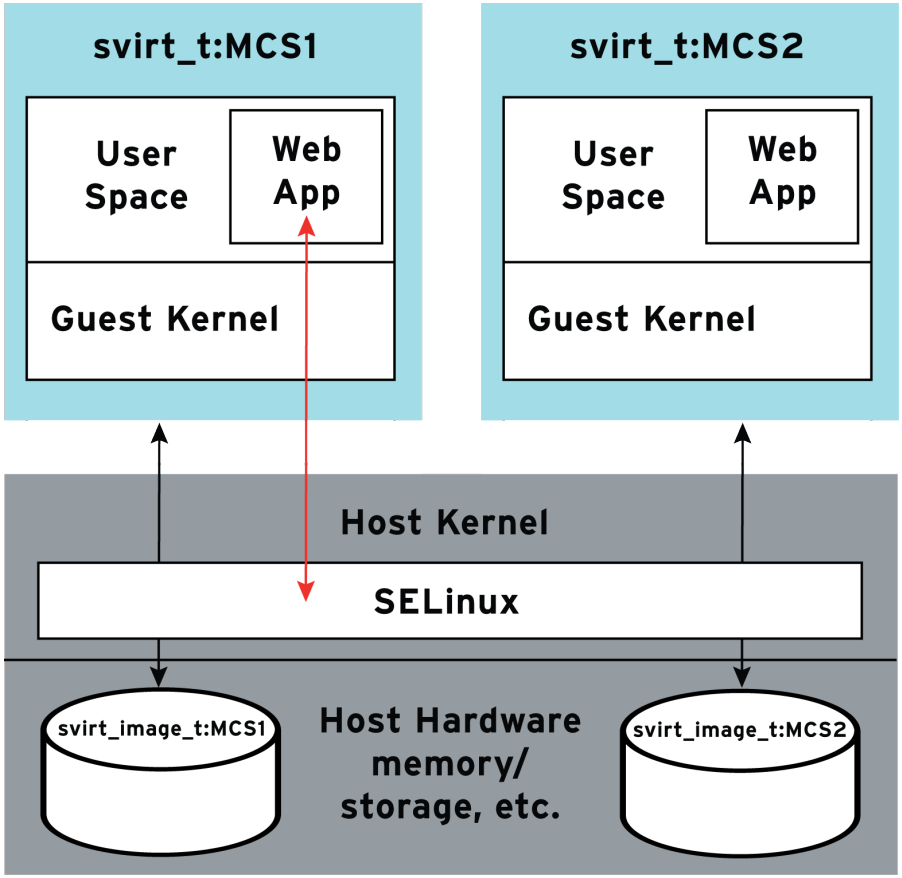
RAS

- timekeeping
- QMP, qdev, VMState, SeaBIOS
- vmchannel
- stable guest HW ABI
  - machine type
  - stable PCI topology
- PCI Device Assignment improvements
- Live migration with large memory guests

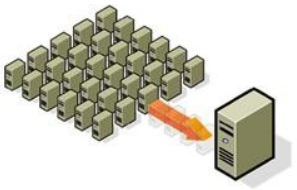
Isolation: sVirt Security



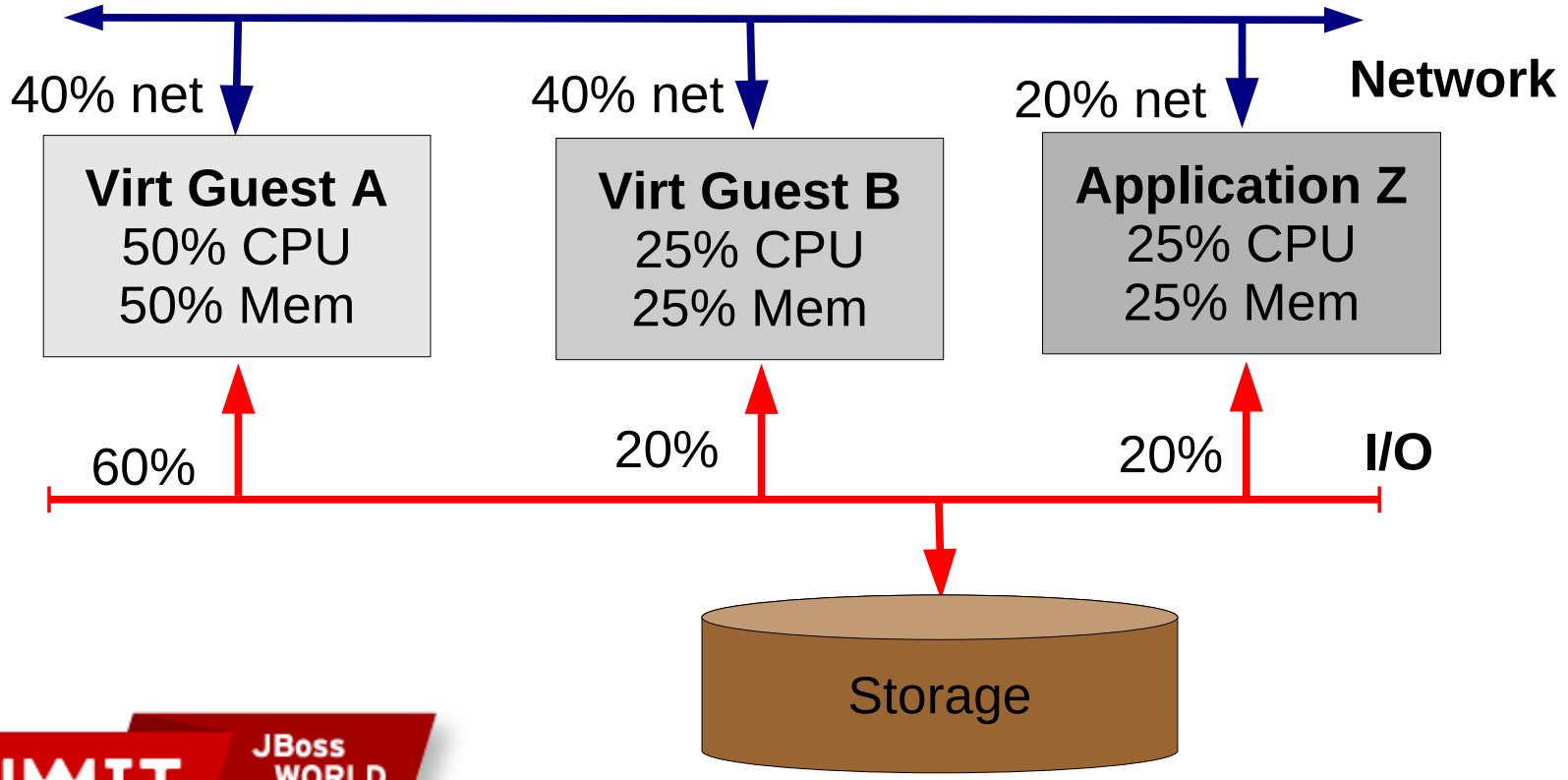
- *Applying security labels to individual guest virtual machines and their resources*
  - Guest Isolation achieved with SELinux Mandatory Access Controls (MAC)
    - Protect against untrusted Guest VM
    - Protect against Host misconfiguration
    - Prevents unauthorized access of Guests/Host
  - Builds on existing, proven security mechanisms & controls



Isolation: Resource Management



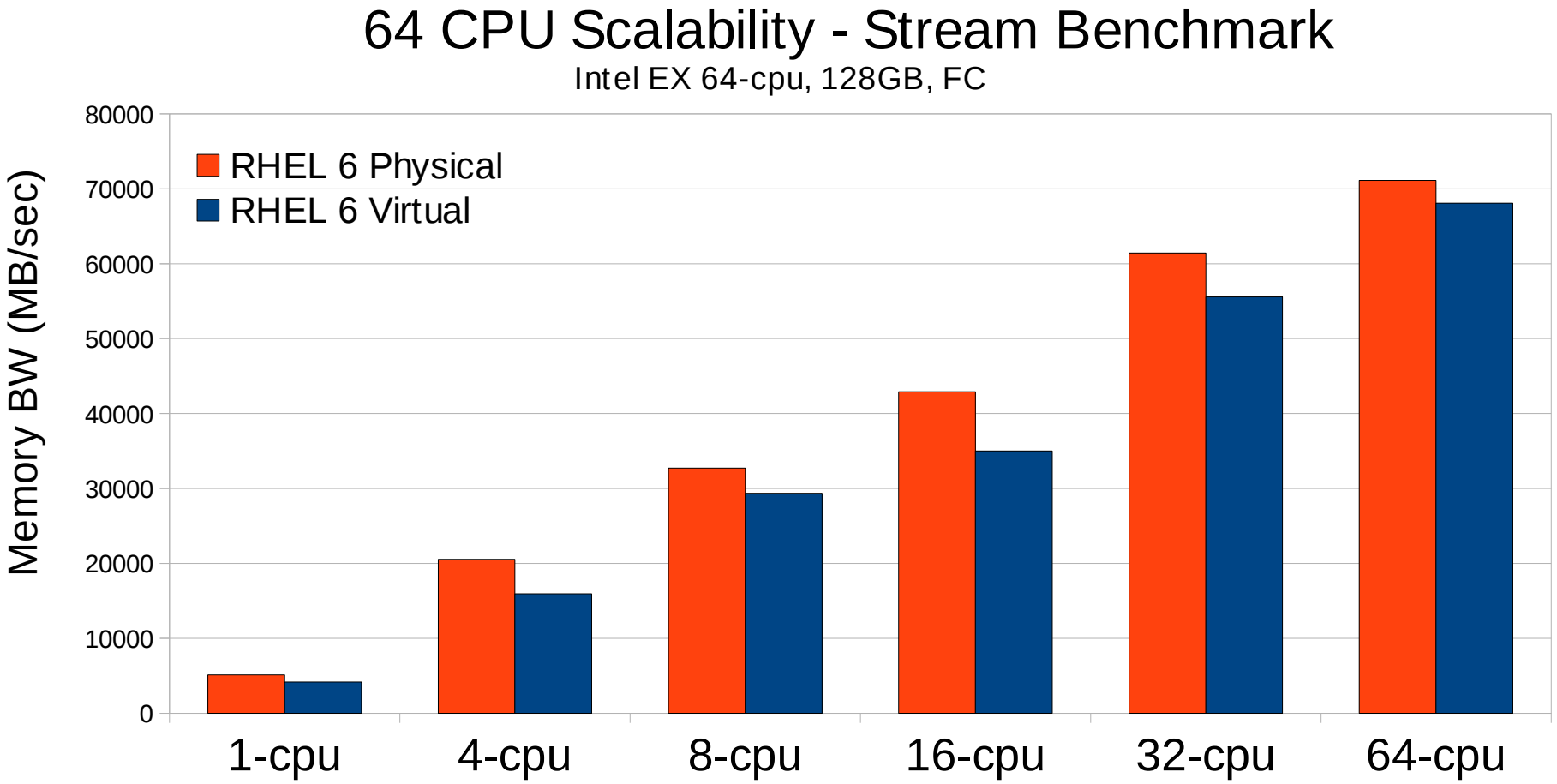
- Control Group (Cgroups) for CPU/Memory/Network/Disk
  - Benefit: guarantee Quality of Service
  - Ideal for: Virtualization/Cloud deployments



KVM: CPU Performance

- Scale to 64 vcpus! And same pcpu scaling as Linux
- Guest spin lock-holder preemption sol'n
- KVM efficiency
  - User return notifiers...get lazier
- x2apic
  - Use MSR access to limit mmio accesses to the irq chip

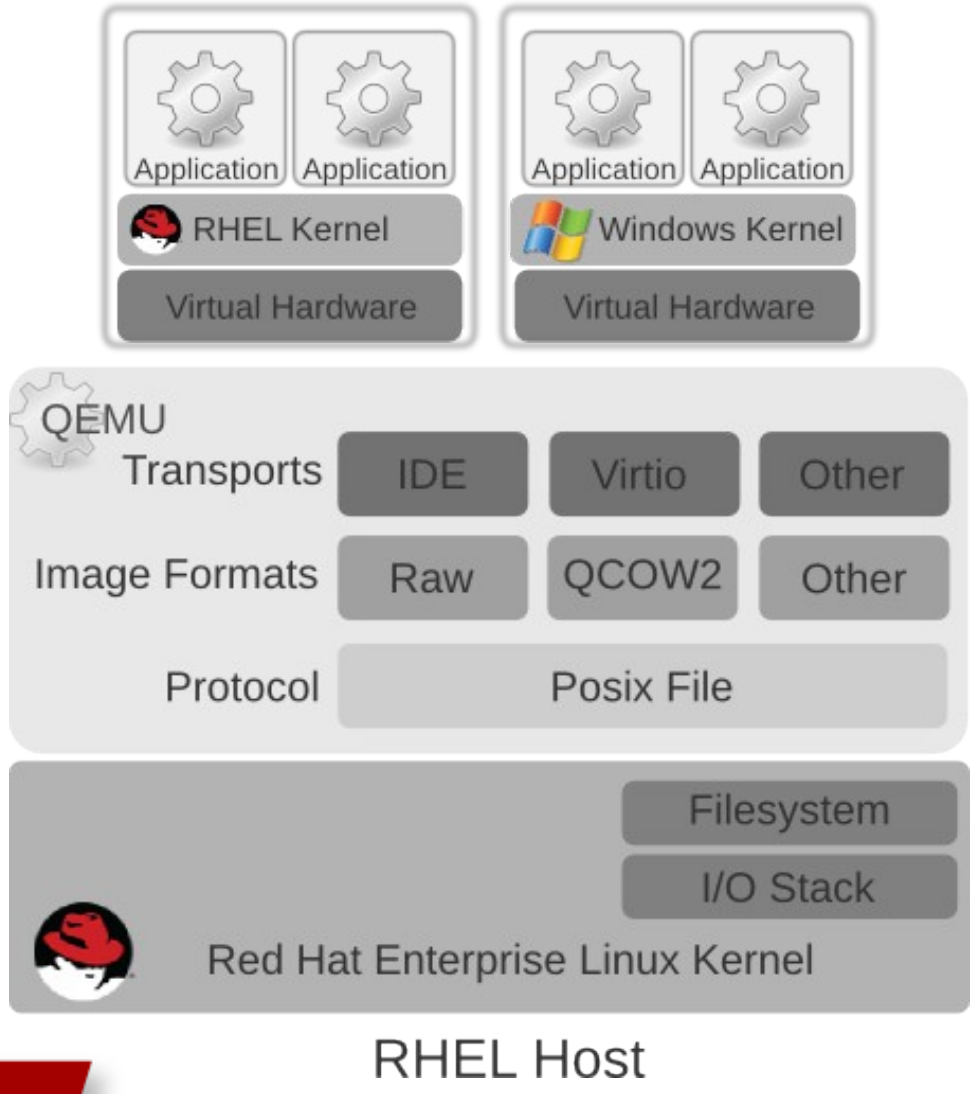
Performance: SMP Scalability



Excellent, linear scalability; minimal virtualization overhead

Note: With this h/w, at 32 CPUs socket bandwidth is saturated

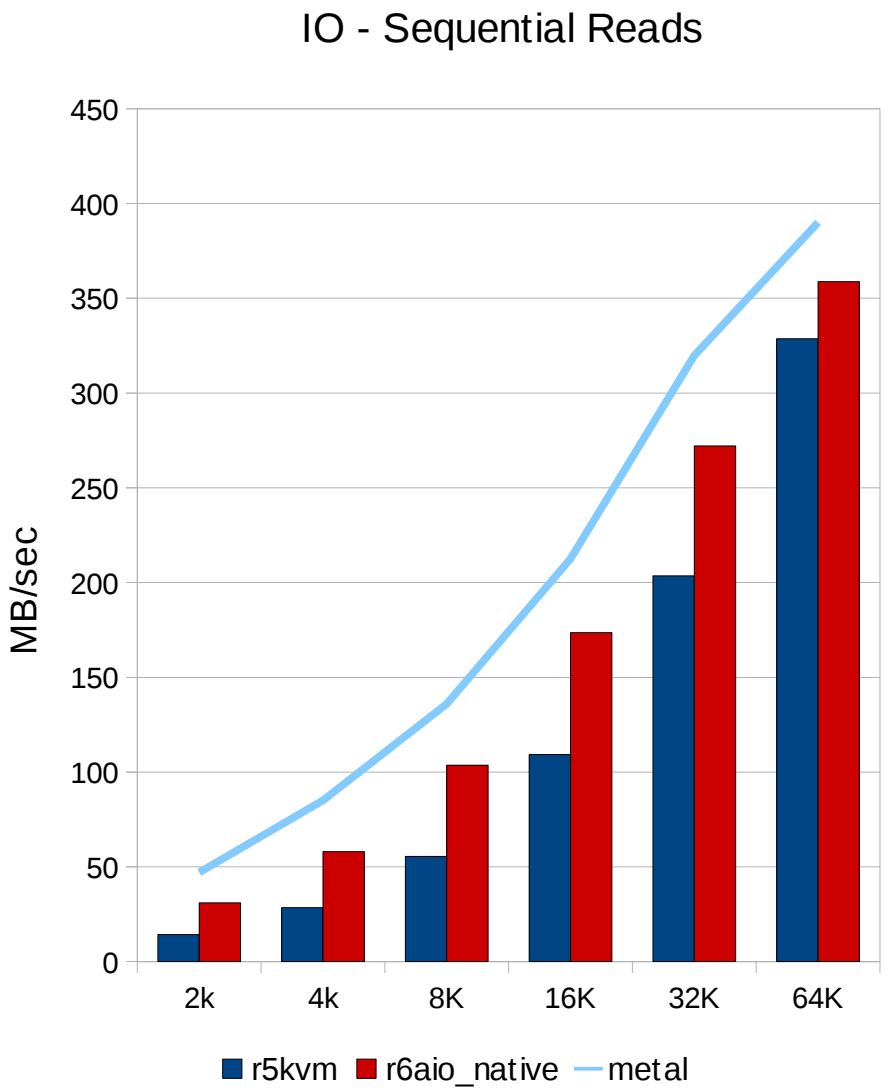
KVM Block Architecture

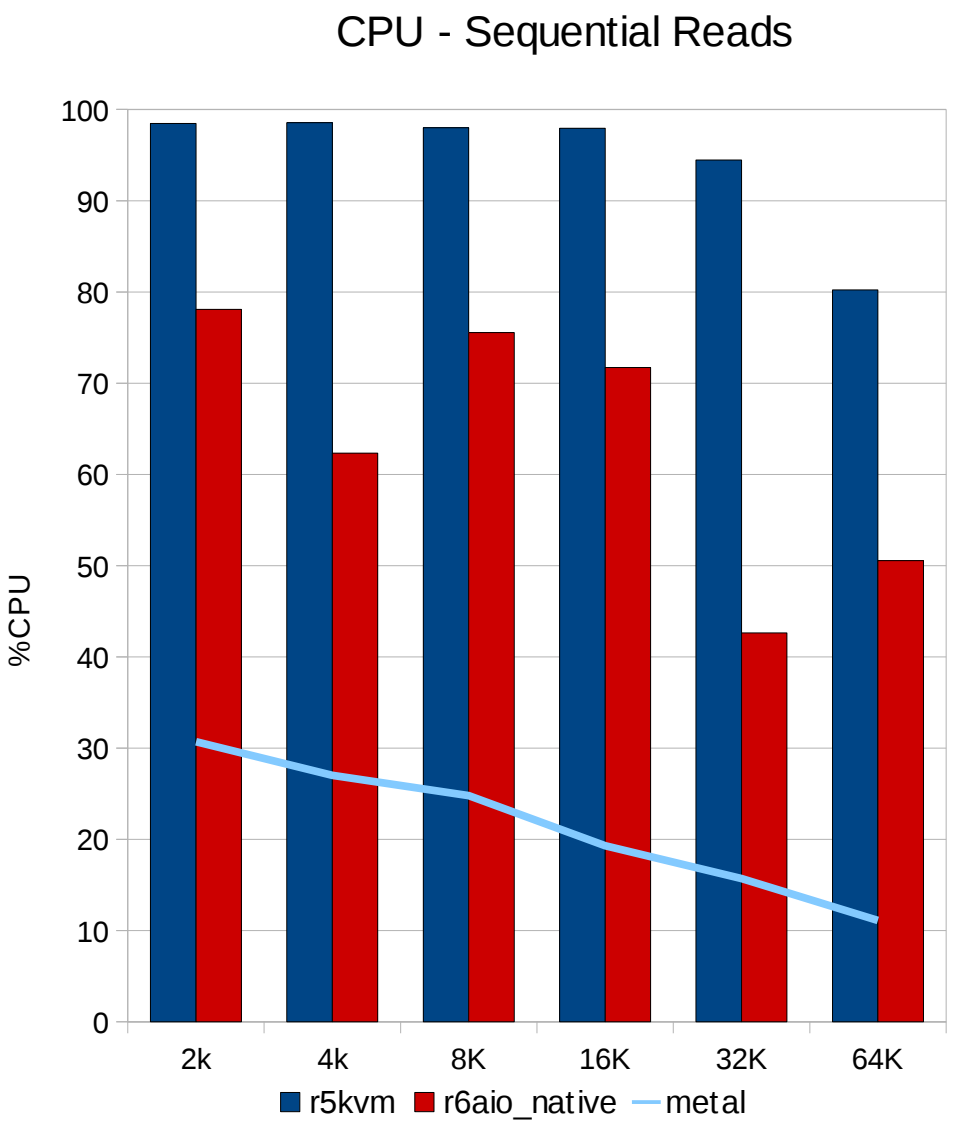


KVM: Block Performance

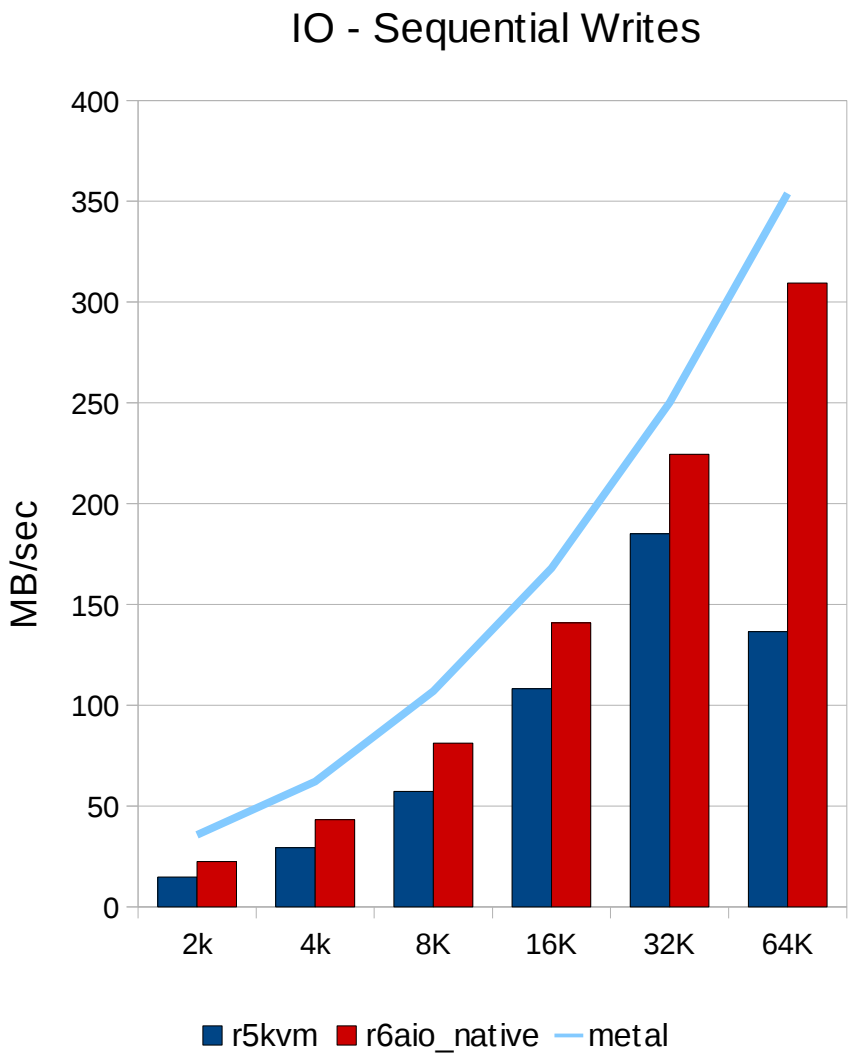
- qemu vectored i/o (preadv/pwritev) and native AIO
- virtio
  - flush + fua
  - MSI support
  - ioeventfd
- qcow2 cache

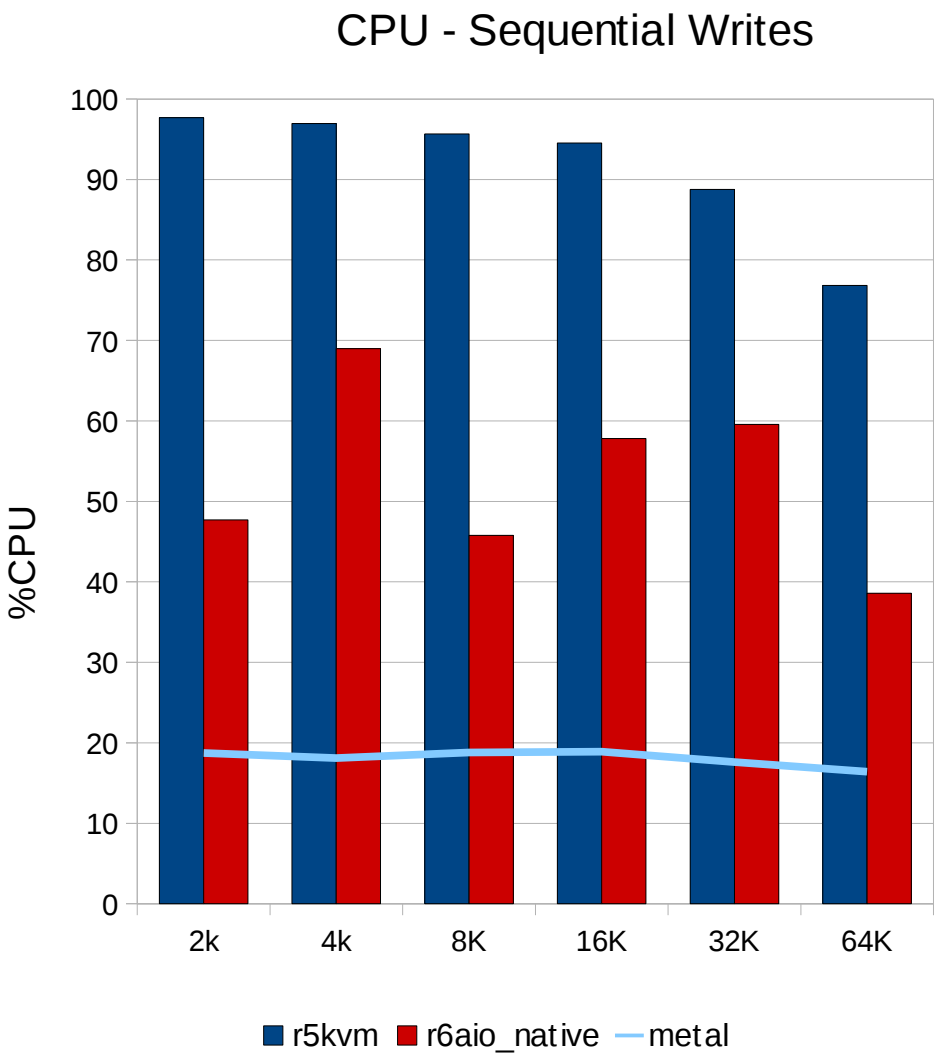
KVM Performance – RHEL6 aio=native Win2k8 Intel 24cpu, 64GB, FC IOmeter



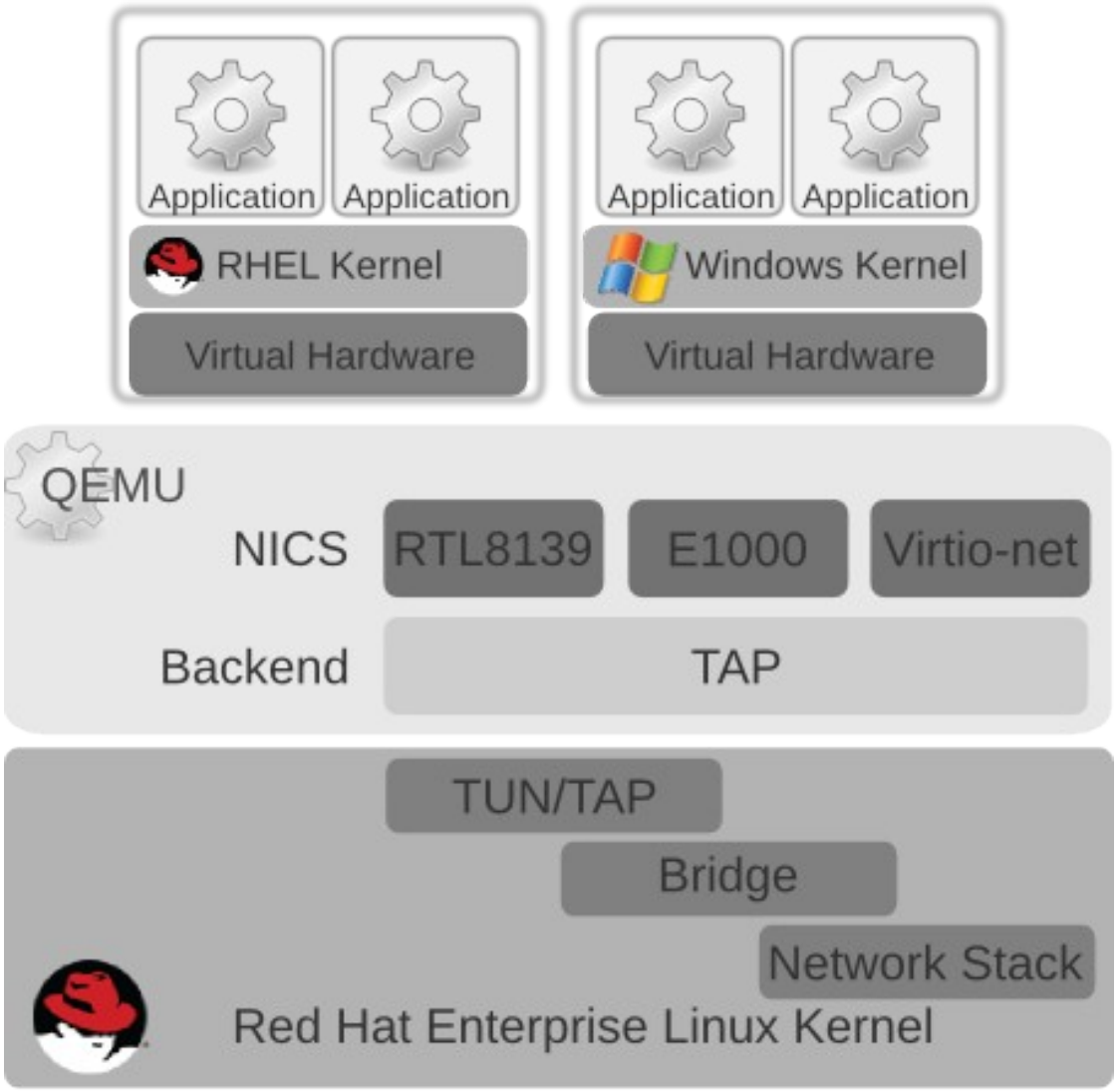


KVM Performance – RHEL6 aio=native Win2k8 Intel 24cpu, 64GB, FC IOmeter





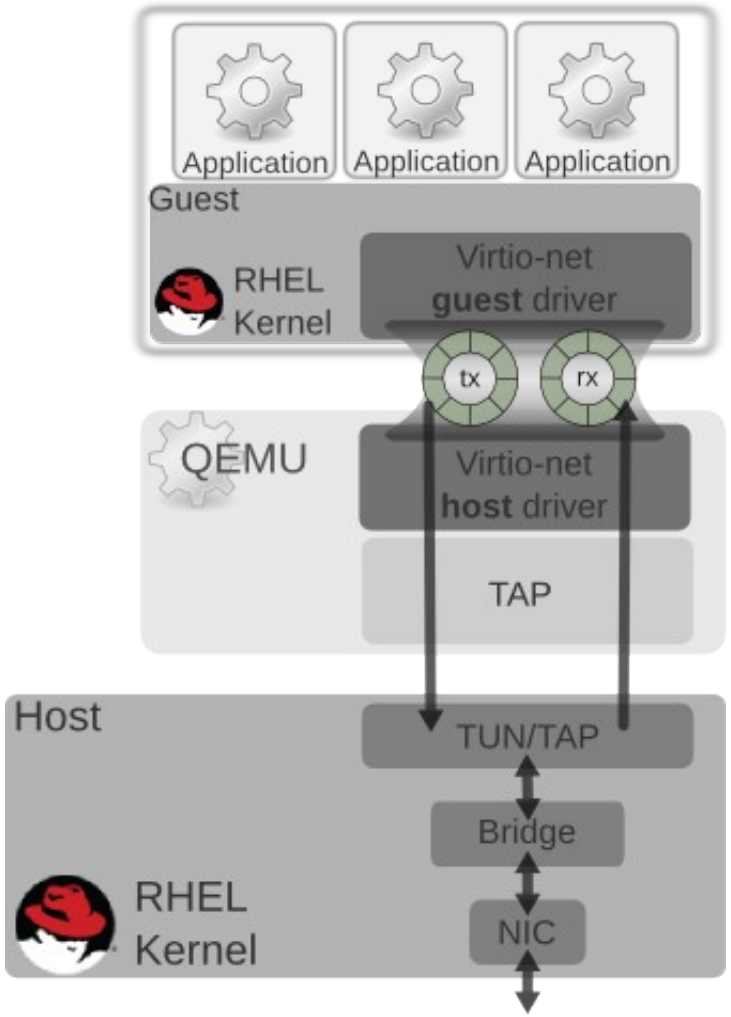
KVM Network Architecture

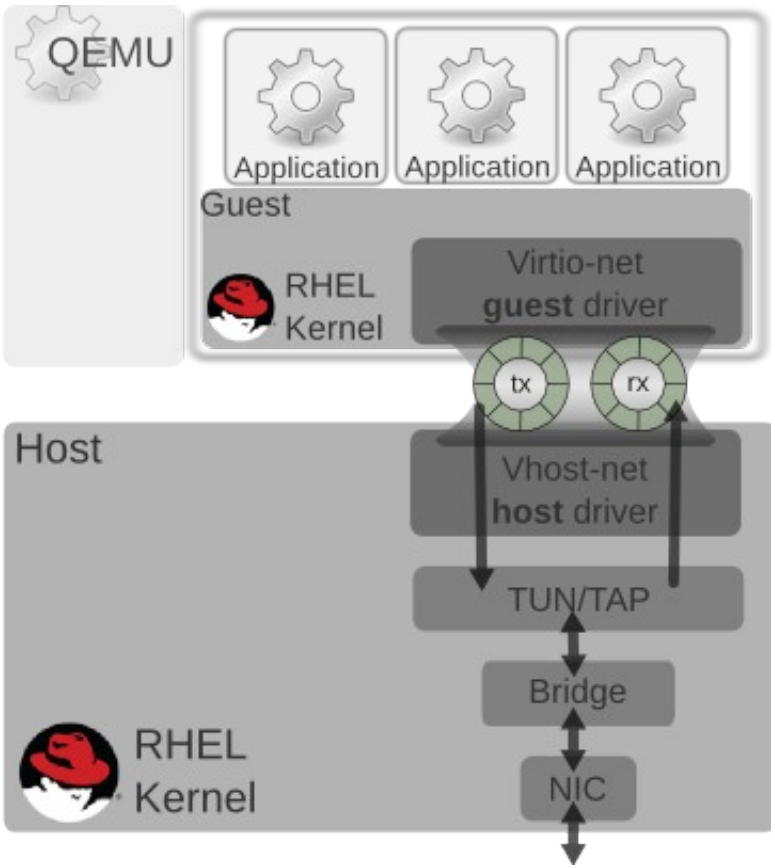


Performance: Networking

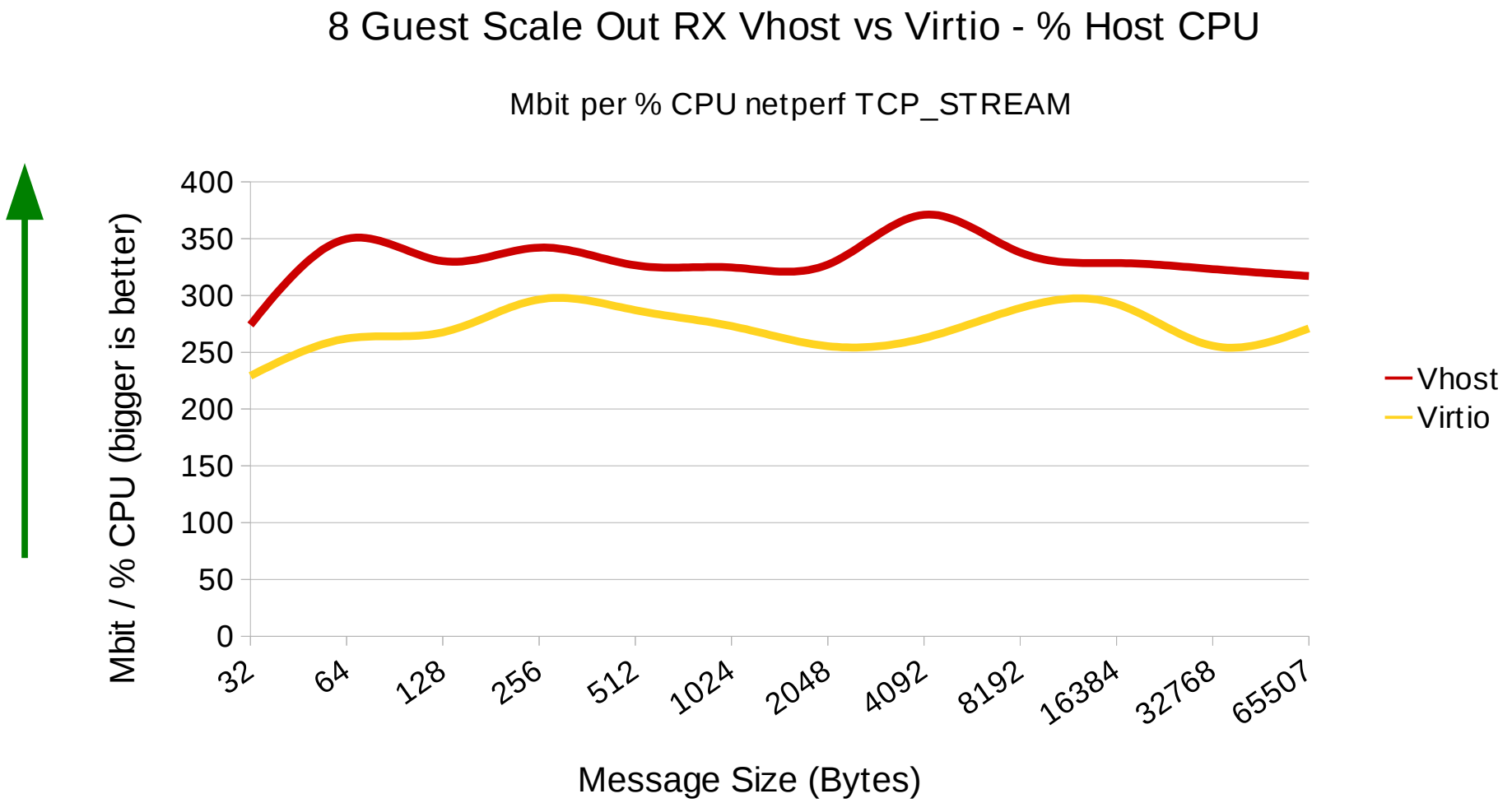
- vhost-net
  - virtio backend in kernel
  - transparent to guest
  - reduce latency
  - better throughput
  - lower CPU consumption
- SR-IOV for near native performance
- VEPA/VNLink

virtio network architecture – detail

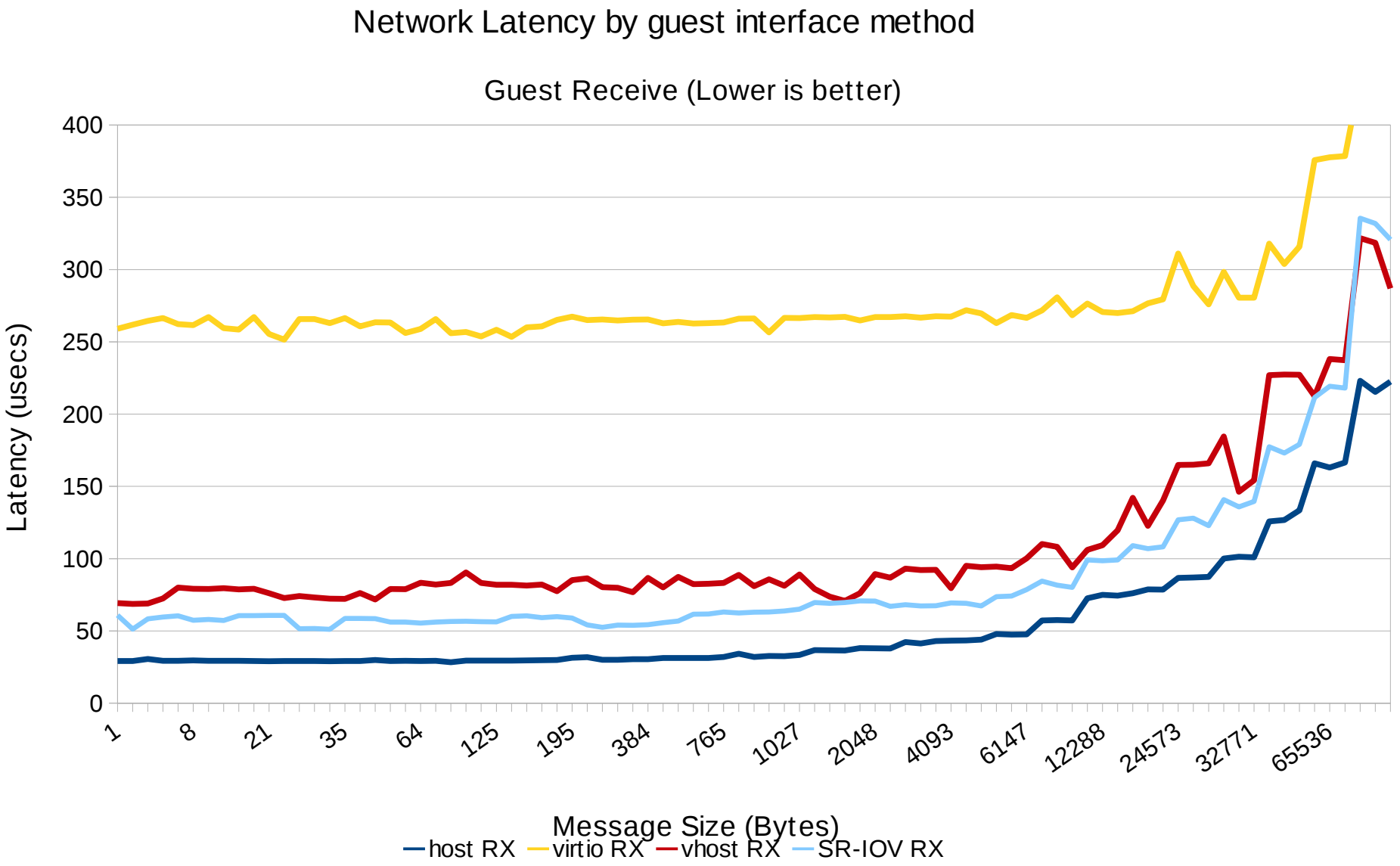




vhost_net Efficiency



Latency comparison – RHEL 6



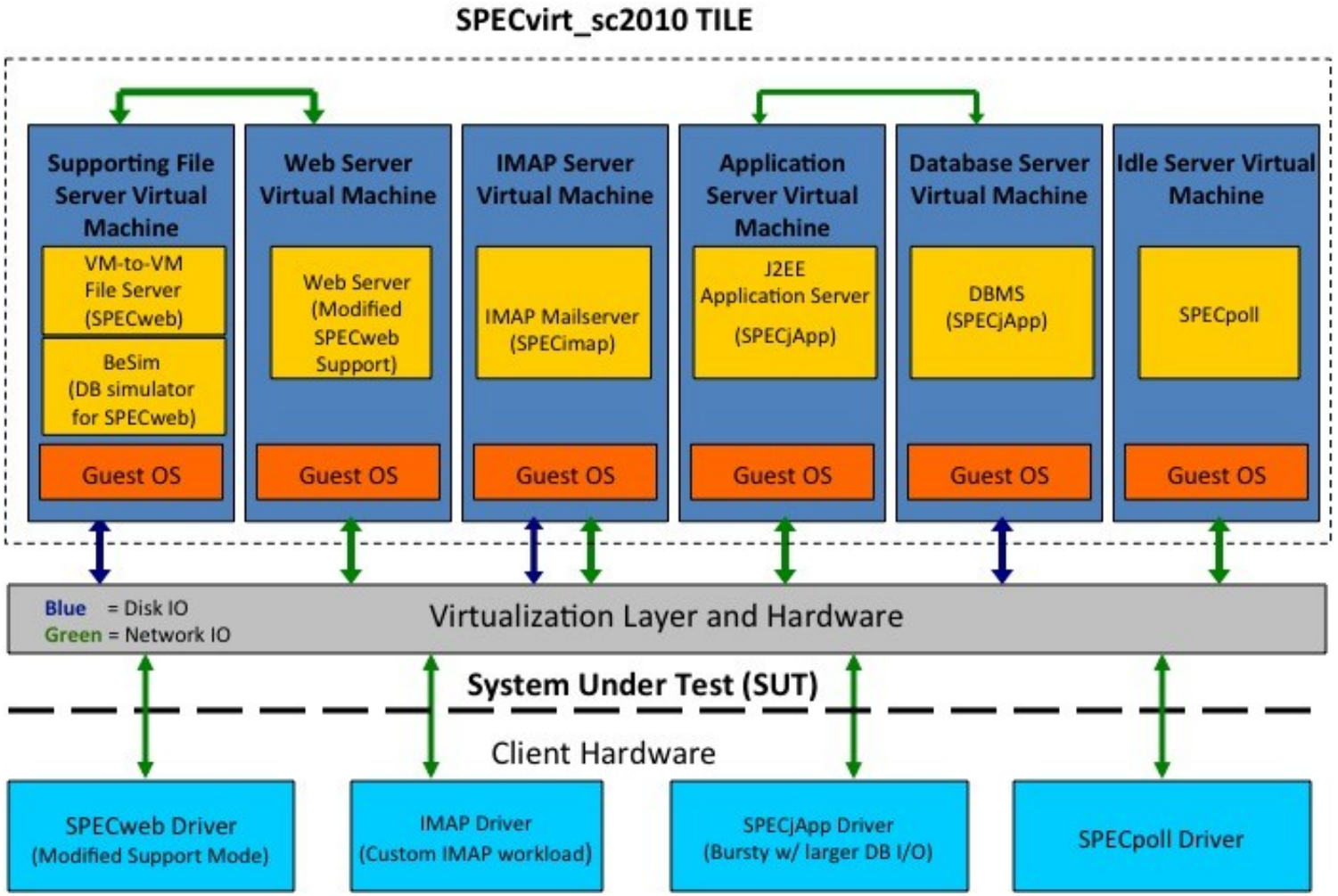
Performance: Memory

- Transparent Huge Pages
  - Benefit for bare metal apps (4.2%)
  - Huge benefit for KVM VMs (25%)
- KSM working with THP

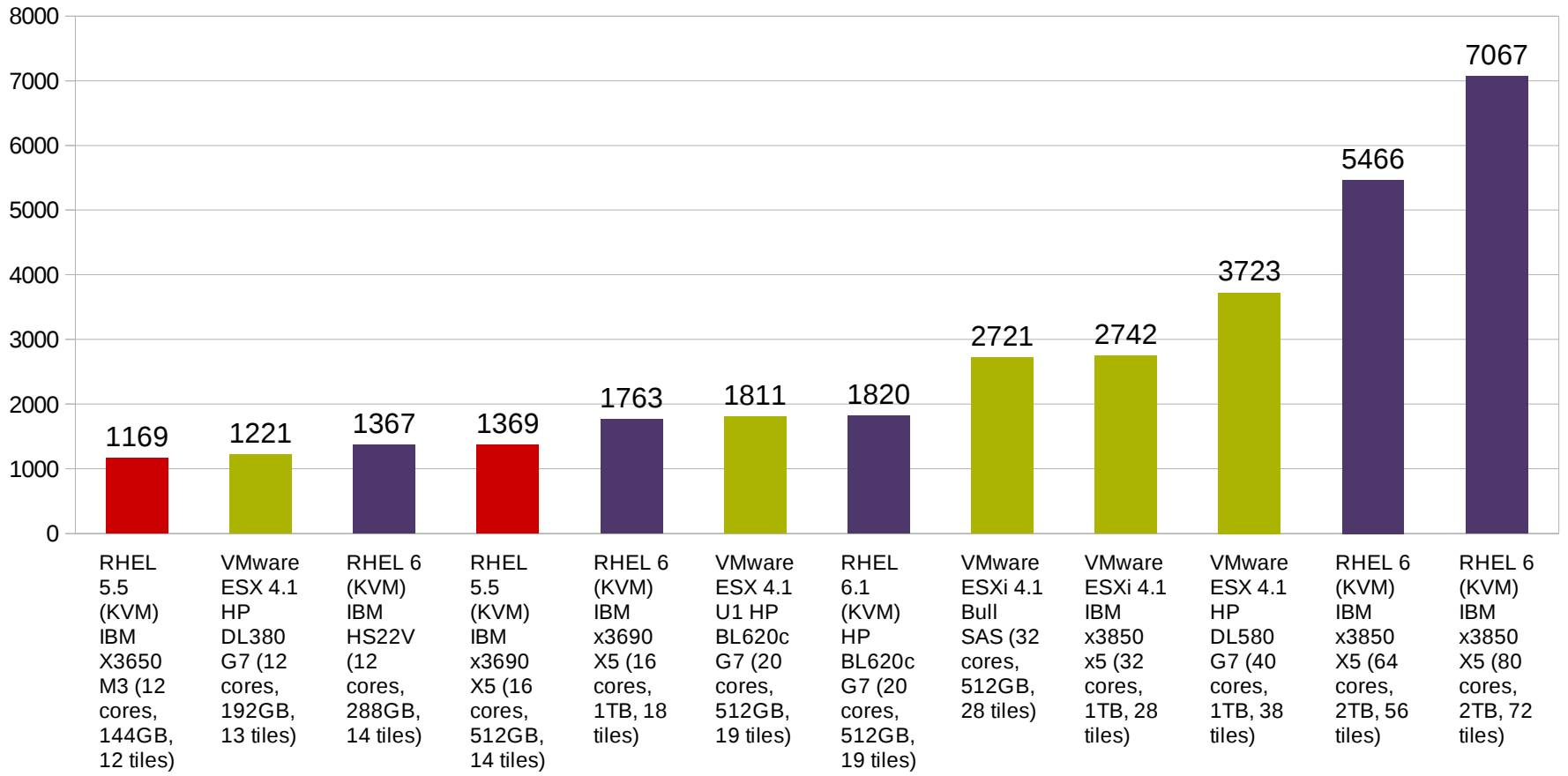
Performance

What does it **add** up to?

SPECvirt


SPECvirt_sc2010 TILE
Supporting File Server Virtual Machine
VM-to-VM File Server (SPECweb)
BeSim (DB simulator for SPECweb)
Guest OS
Web Server Virtual Machine
Web Server (Modified SPECweb Support)
Guest OS
IMAP Server Virtual Machine
IMAP Mailserver (SPECimap)
Guest OS
Application Server Virtual Machine
J2EE Application Server (SPECjApp)
Guest OS
Database Server Virtual Machine
DBMS (SPECjApp)
Guest OS
Idle Server Virtual Machine
SPECpoll
Guest OS
Blue = Disk IO
Green = Network IO
Virtualization Layer and Hardware
System Under Test (SUT)
Client Hardware
SPECweb Driver (Modified Support Mode)
IMAP Driver (Custom IMAP workload)
SPECjApp Driver (Bursty w/ larger DB I/O)
SPECpoll Driver

SPECvirt_sc2010 results*



*As of May 4, 2011. See http://www.spec.org/virt_sc2010 for details

Agenda

Virtualization Overview

Foundation for Cloud

Roadmap

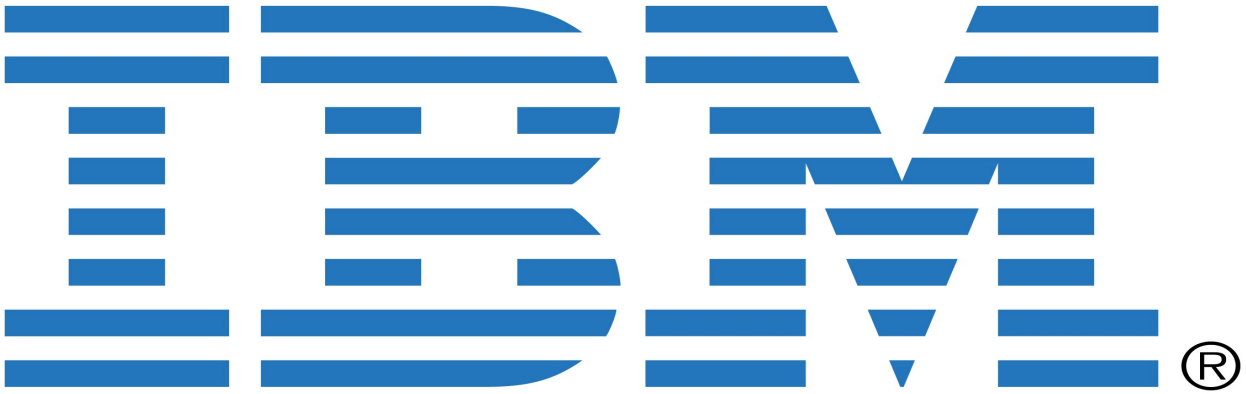
Cloud

- Computing as a Service (Anything as a Service)
- Utility model
- Self provisioning
- Dynamic resources

KVM in Cloud

- Increase density
- Improve efficiency
- Multi-tenancy (security, isolation)

KVM Is The Cloud



Agenda

Virtualization Overview

Foundation for Cloud

Roadmap

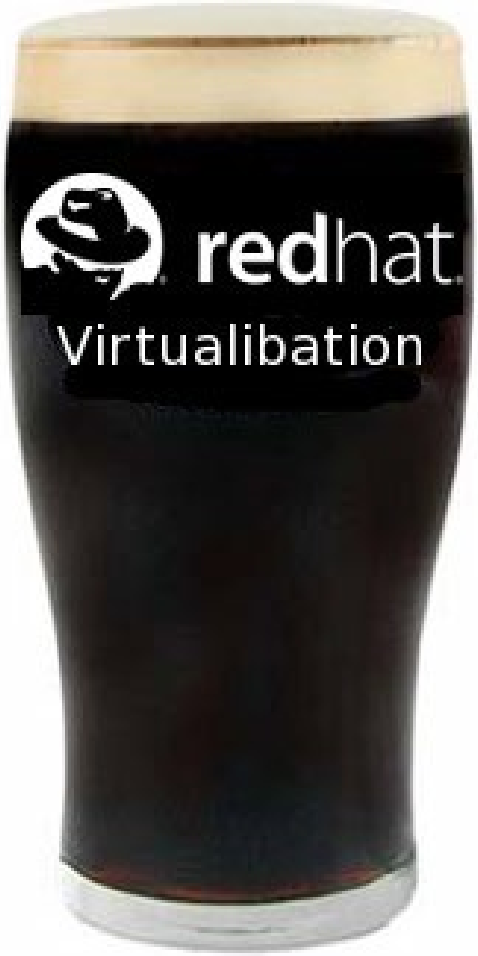
Roadmap

- Scaling and performance
- NUMA
- Lock holder preemption
- Zero copy networking
- Multiqueue virtio-net
- Debugging and profiling
- Further cgroup integration

- QCOW2 performance
- Multi-tenant network isolation
- Memory management (async page faults, hinting, compcache)

Weather forecast...

Cloudy
with a chance of
total world domination


redhat.
Virtualibation

LIKE US ON FACEBOOK

www.facebook.com/redhatinc

FOLLOW US ON TWITTER

www.twitter.com/redhatsummit

TWEET ABOUT IT

#redhat

READ THE BLOG

summitblog.redhat.com

GIVE US FEEDBACK

www.redhat.com/summit/survey